

## 算数問題文の解析とその知識表現

笹田 国博\* 小高 知宏\*\* 高濱 徹行† 小倉 久和\*\*

### Analysis of the Arithmetical Word Problems and Representation of the Knowledge

Kunihiro SASADA\*, Tomohiro ODAKA\*\*, Tetsuyuki TAKAHAMA†, and Hisakazu OGURA\*\*

(Received Feb. 28, 1994)

In this paper, we investigate semantic analysis in natural language processing. There are various rules and expressions for different purposes or fields even if one uses a single language. However, if a situation is restricted, we can treat the meanings of the sentences by the use of a characteristic of the representations in language. We choose arithmetical word problems in Japanese for elementary school children as an object of this study. These problems are comparatively easy and simple for automated analyzing. We use Common Extended Self-contained Prolog(CESP) to analyze the Japanese sentences of arithmetical word problems and to represent the semantics.

#### 1 はじめに

本研究では、自然言語処理の中の意味解析に着目する。同じ言語でも、使う分野や目的によって使い方や表現方法が異なったり分野特有の表現や規則が存在する。われわれは、言語の用いられる局面を限定すれば、その言語特有の特徴を利用し、言語というものの意味理解を図ることができるのではないかと考えた。そこで比較的簡単な文で構成されている小学校低学年の算数における日本語文章題を対象として選び、CESPによる日本語文の意味表現と理解の試みを行った。ここでは、算数文章題の知識処理の特徴と、CESPによる小学校低学年の算数文章題の文解析と意味理解の試みを紹介する。

算数問題の世界では、我々の日常生活の出来事が広く扱われている。したがって問題の中で述べられる様子や変化は問題を解く人が状況をイメージしやすいものが選ばれている。対象を小学校低学年においた問題では、小学校低学年の児童の目から見た日常の世界に関する表現がされている。

問題によっては、図や挿絵などを利用して状況理解を図るものもあるが、本研究では算数の文章題を利用した言語処理を目標としているので、これらの問題は扱わない。以降で述べる算数問題とは、挿絵などの視覚的な要素を必要としない文章のみで判断が行われる算数問題のことを指す。

小学校低学年の算数問題では、文章から状況を読み取り、それらの関係を表現する式を導き出すことの能力を求めている。したがって、文章の意味理解をさまたげるようなまぎらわしい表現や難しい表現は存在しない。また算数に関する文章であることから、必ず数に関することが質問されている。したがって算数問題の内容は必ず数の状況変化に関するものが示されており、必ず1ひき、2ひきといったような数の単位が存在している。ただごく稀に単位を必要としないものもある。この単位のこ

\*大学院情報工学専攻

\*\*工学部情報工学科

†情報処理センター

とを助数詞と呼ぶ。また、算数問題は物の様子や状態の変化を問うものであるので、問題文中には必ずその変化の対象となるものが存在する。

状態の変化の様子や内容を表す要素として、述語の存在が挙げられる。述語はどの言語にも存在し、文章の意味を決定するのに大きな役割を果たす。算数問題においては、述語は計算方法を決定する上で重要な役割を果たすものである。小学校低学年を対象とした問題文に出てくる述語は我々の日常生活において使われるものを多く集めたものであり、それらより計算方法を決定するためには述語に関する多くの知識を必要とする。

そこで本研究では、日本語の膠着性を利用した述語中心の解析による小学校低学年の算数問題の解析を行なう。この際、われわれが普段算数の文章題を解く際の考え方を基に、文章題をいくつかのタイプに分類し、その分類したタイプに問題文をユニフィケーションすることによって文章題を解いていく。この分類には、教科書や参考書に見られるような分類ではなく、知識表現の観点から見た文章題の分類を行なった。

本研究では意味解析に重点をおき、他の処理を出来る限り軽減するために、字句・構文解析は殆ど行なわない。このため、対象とする問題文は、あらかじめひらがな表記に直しておき、ある程度の文節毎に空白により区切り文字を入れておくこととした。さらに対象は1回のみの加減算を行なう問題とし、繰り返し演算を要求するものや複数の解答を要求する問題は今後の課題とした。

## 2 算数問題文の特徴と、それによる分類および表現

### 2.1 算数問題文の特徴

算数問題の文章題は事物の様子や状態の変化を問うものであり、必ず変化の対象となる対象物及びその数の状況を示す数詞が存在する。また意味理解を行う上で必要な情報も基本的にはこの2つと述語だけである。たとえば、

ひろさんは えほんを 8 さつ もっています。  
おばさんに 2 さつ もらいました。  
えほんは ぜんぶで なんさつに なりますか。

といった問題について、解答するのに必要な情報は

えほんを 8 さつ もつ。  
(えほんを) 2 さつ もらう。  
えほんは ぜんぶで なんさつに なるか。

であり、この情報さえあれば問題は解ける。ひろさん、おばさんといった情報は無視しても良く、実際我々が算数問題を解く際には、このような情報抽出を行っていると思われる。また、対象物は問題によっては1文の中に複数あったり、文脈から判断出来る場合は省略されている可能性があり、その時には推論により対象物の同定を行なうことが必要となってくる。

その他に計算方法の決定に重要な意味を持つ算数の文章題特有の要素として、

まるい さらが 2 まい あります。  
しかくい さらが 6 まい あります。  
ちがいは なんまいでしょう。

のような問題の中の、質問文の中の「ちがいは」という要素がある。この要素が「みんなで」となることによって答の計算方法は全く違ったものとなる。この計算方法を最終的に決定する要素を以降キーワードと呼ぶ。本研究では、対象物、数、助数詞、述語、キーワードの情報を基に解析を行う。

算数問題にはしばしば、文章中には現れないが我々が意識せずに使用している常識を必要とするものがある。例えば、次の問題の場合1文目の「おりがみ」が2文目では「つる」に変化している。この関係が読みとれなければ問題を正しく解答することは出来ない。

おりがみが 10まい あります。  
つるを 4つ おりました。  
おりがみは なんまい のこって いますか。

また次の問題文では、「こども」は「おとこのこ」と「おんなのこ」の2つに排他的に分類されるという知識を必要とする。

こどもが 15 にん あそんで います。  
その うち 8 にんは おんなのこです。  
おとこのこは なんにん でしょう。

このように問題によっては、表現が違っているのに同じ物を指している場合や、その状態を経験した人にしかわからないような物質の変化の様子が既知のこととして明記されている問題文が存在する。このような問題文は特に小学校低学年の算数問題に多く存在する。これは、小学校低学年の算数問題は児童に対し、算数に関する知識だけでなく、初等教育における文章の推論の知識の形成にも関与しているためであると考えられる。このことは算数問題の特徴の1つとして考慮しなくてはならない問題であると考えられる。

本研究で開発・作成してきたシステムでは、現在のところ、これらの常識知識ともいえる知識については辞書の拡張と推論の例外的処理で対処している。

次の文章のように、

ひごいが 4 ひき いる。  
まごいが 2 ひき いる。  
あわせて なんびきでしょう。

1文目で「ひごい」、2文目で「まごい」とそれぞれ違った対象物が存在している時でも、助数詞などを手がかりとして、共通の特徴が抽出できれば、それをもとに解析は可能である。この場合であると、「ひき」という助数詞が共通であることと、「ひごい」、「まごい」が持つ属性から、質問文で聞いている対象物を「さかな」と推定することができ、答を導き出すことが可能である。そこで、本研究では、システムが持つ辞書の各項目にそれぞれの単語が持つ属性を持たせておき、解析の際には一見異なる対象物の場合でも、属性が同じであれば、その属性を基に解析を進める。

しかし、先程の「おりがみ」が「つる」に変化する例のような、対象物が全く違う表現へと変化してしまう問題には、この方法では解決できない。「おりがみ」の属性と「つる」の属性はやはり普通の観点では違うものである。この2つの対象物を結びつけるには、算数問題の世界からさらに状況を絞り込んでその問題中の世界を構成し、そこで動的に推論知識を構成することが必要であると思われる。本システムでは、対象物がこのような変化をする問題文については解析できない。

また、「こども」が「おとこのこ」と「おんなのこ」に排他的に分類されるという知識に関しては、現在システムでは「こども」という対象物が出てきた際に推論の例外的処理として、「おとこのこ」、「おんなのこ」の2つの対象物に分割し、解析を進める。現在のところ、この例外的処理は「こども」に対してのみ行なっている。

## 2.2 算数問題文の文構造による類型分類

算数問題文の形にも様々なものがあるが、小学校低学年を対象としたものでは、3文程度で構成され、問題文中に現われる2つの数字を使って計算するものが殆どである。そして、その問題文の構成は、はじめの2文である場面での状況やその変化の様子を記し、最後の文でなんらかの数を聞く、という形となっている。3文でない問題も、その殆どが2文を複文として1文に記述したもの、文章の数を増やし同じような計算を繰り返すものなどであり、基本的には同型と考えられる。

3文から構成される問題は意味的な文構造から次のように、4つに分類して考えることが出来る。

### (1) タイプ A：対象物の数を聞く問題

条件文1で物事の状態を表現し、条件文2でその変化の様子を示し、その変化による結果を質問文で尋ねる問題である。したがって条件文1では状態を表す述語、条件文2では変化の様子を表す述語、質問文ではその結果を聞く述語という組み合わせになっている。例えば以下のような問題文である。

がようしが 6まい ある。  
2まい つかう。  
なんまい のこるでしょう。

この文では条件文2と質問文に対象物が示されていないが、対象物が同じものの時はこのように省略されることがよくある。我々は普段これを文脈から推論することによって他の文中より抽出して行く。本研究においても、これらの省略された対象物は他の文章中より見つけだすことによって解析を行う。

### (2) タイプ B：条件文2が比較を表す問題

条件文2に2つの対象物の比較を表現している文章が入っている問題である。この場合条件文2には、比較を表現する述語と比較される被対象物と対象物が含まれる。例えば以下のような問題である。

すいそうに あかい きんぎょが 15 ひき いる。  
くろい きんぎょは あかい きんぎょより 6 ひき すくないです。  
くろい きんぎょは なんびき いるか。

この問題のように比較の問題では、同じ対象物をその様子で区別する表現がよく使われる。このため対象物には形容詞が付く。この形容詞は文章中の対象物の違いを示す要素であるので重要である。

条件文1や質問文の構成はタイプAの問題と同じである。この問題では条件文2の文章より2つの対象物の違いを理解し、条件文1の状態と照らしあわせて質問されている対象物の状態を推論しなければいけない。したがって条件文1と質問文に現れる対象物と条件文2に現れる2つの対象物はそれぞれ対応していなければならない。

### (3) タイプ C：質問文が比較を表す問題

これはタイプBの問題と良く似ているが質問文に比較を表す文章がある問題である。比較する被対象物、対象物がそれぞれの条件文に記されていて、質問文でそれらの比較を行う。例えば以下のような問題である。

おとこのこが 34 にん いる。  
おんなのこが 20 にん いる。  
おとこのこは なんにん おおいでしょう。

このような、対象物をその様子で区別する問題文の場合では、同じ単語の重複をさけるために、質問文中では対象物の省略や代名詞が使用される。したがって解析をする際には、これらをおぎなうための推論を行わなければならない。

#### (4) タイプ D：条件文 1 に 2 つの対象物がある問題

条件文 1 の文章に 2 つの対象物が存在し、それらを合わせた様子が記述されており、条件文 2 で、それら 2 つの対象物のうちの片方の様子が書かれており、質問文で残りの対象物の状態を聞く問題である。例えば以下のような問題である。

あかと あおの いろがみが あわせて 16 まい ある。  
 その うち あおは 7 まい です。  
 あかは なんまい でしょう。

この問題のように対象物 (いろがみ) を省略し、その様子 (あか、あお) で対象物を表現することも多い。この場合も今までの場合と同じように、前の文章から対象物を推論することによって解析を行う。

### 2.3 述語表現による算数問題文の表現

算数問題の内部表現は CESP の述語で表現する。文章の述語や対象物、数詞などは述語の引数として、関数の形式で表す。システムは入力された問題文から必要なものを探索・推論し、この表現を構成する。

述語 (引数 1、引数 2、.....、引数 N);

以下に問題のタイプごとの内部表現方法をまとめる。

#### 1. タイプ A

```
条件文 1(Class,jutugo(述語),object(修飾詞,対象物),suu(数,助数詞));
条件文 2(Class,jutugo(述語),object(修飾詞,対象物),suu(数,助数詞));
質問文(keyword(キーワード),jutugo(述語),object(修飾詞,対象物),
suusi(助数詞));
```

修飾詞や質問文のキーワードはあってもなくても良い。

#### 2. タイプ B

```
条件文 1(Class,jutugo(述語),object(修飾詞,対象物),suu(数,助数詞));
条件文 2(Class,jutugo(述語),object1(修飾詞,対象物),
object2(修飾詞,対象物),suu(数,助数詞));
質問文(jutugo(述語),object(修飾詞,対象物),suusi(助数詞));
```

条件文 2 には比較を表すために 2 つの対象物がある。

#### 3. タイプ C

条件文 1(Class,jutugo(述語),object(修飾詞,対象物),suu(数,助数詞));  
 条件文 2(Class,jutugo(述語),object(修飾詞,対象物),suu(数,助数詞));  
 質問文(jutugo(述語),object1(修飾詞,対象物),  
           object2(修飾詞,対象物),suusi(助数詞));

質問文が比較文になっており、対象物が2つある。

#### 4. タイプ D

条件文 1(Class,jutugo(述語),object1(修飾詞,対象物),  
   object2(修飾詞,対象物),suu(数,助数詞));  
 条件文 2(Class,jutugo(述語),object(修飾詞,対象物),suu(数,助数詞));  
 質問文(jutugo(述語),object(修飾詞,対象物),suusi(助数詞));

条件文 1 において 2 つの対象物が必要である。

### 3 算数文章題理解応答システムの概要

#### 3.1 算数文章題理解応答システム

前章で述べた内部表現化、推論処理を行なう算数文章題理解応答システムを CESP 処理系を用いて作成した。

システムの構成、流れは図 1 及び図 2 のようになっている。

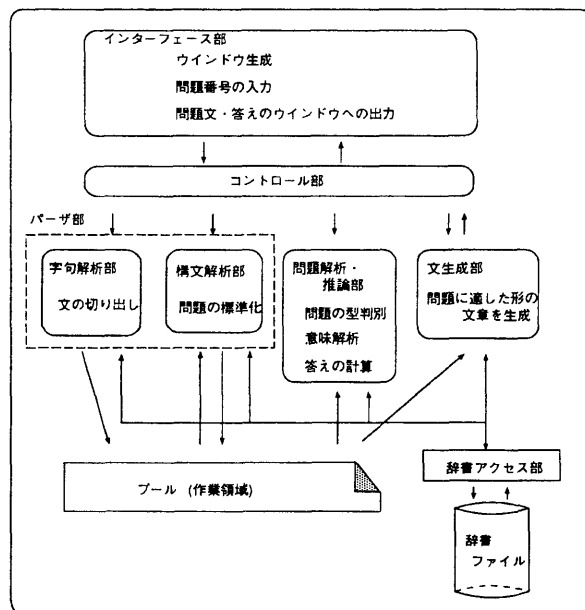


図 1: システム構成

本システムを設計する際に、

- CESP のウィンドウ・インターフェース・ライブラリを利用し、使いやすいユーザ・インターフェースを提供する。
- システム内部で扱う知識(辞書)はシステム動作時にハッシュを利用して管理するものとし、知識を記述したファイル(辞書ファイル)は平文の形で記述する。
- 各処理部がやりとりするデータを、作業領域であるプールで管理することによって、各処理部の依存関係を出来るだけ無くし、CESP のオブジェクト機構を利用したモジュール分けを実現する。

といった方針をたてた。

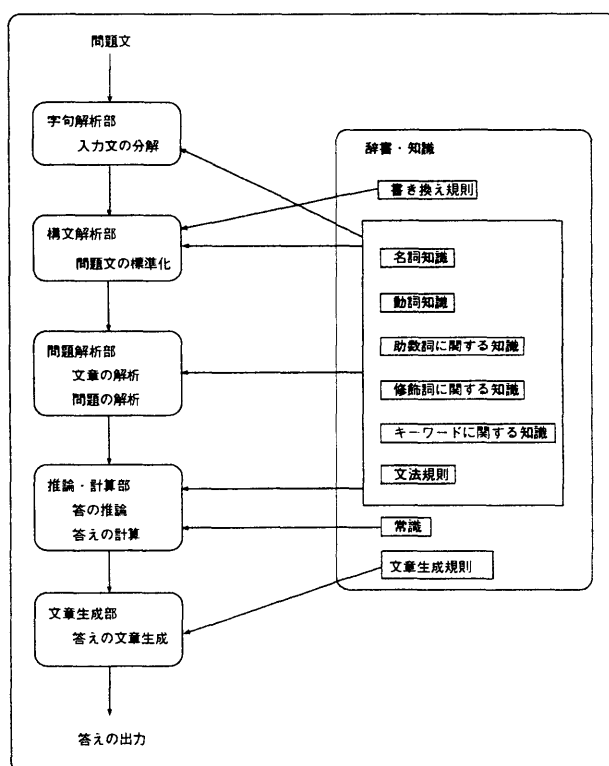


図 2: システムの流れ

ユーザとの対応は全てインターフェース部で行い、インターフェース部により作成されたウィンドウを通じて行われる。また、文章生成部で解答の文章が生成されると、その解答を出力用のウィンドウに出力する。

インターフェース部でユーザより問題番号が入力がされると、インターフェース部は対応する問題をファイルから読み込み、パーザ部に問題文と問題番号を渡し処理を依頼する。

パーザ部では問題文のリスト化及び字句の置き換えが行われる。パーザ部は処理結果を共通の作業領域であるプールへ格納し、処理を問題解析部へ移す。

問題解析部では入力された問題文がどのタイプの問題であるか判断し、判断されたタイプに従った内部表現を行う。

推論・計算部では問題解析部より得られた内部表現より計算方法を推論し、実際の計算を行う。そこで得られた答より文章生成部で答の文章を生成する。

そして、各々の処理部の結果は、システムの共通作業領域であるプールへ格納され、後続の処理部が参照、更新していく。このプールへの各処理部のアクセスは CESP が提供するプール管理ライブラリを利用した。また、このプール自体も CESP が提供しているプール管理ライブラリの中のクラス list を利用する。

以上のような各処理部は、CESP におけるオブジェクトとして構成している。各オブジェクトは図 3 のように処理単位ごとに分けられたメソッド群であるクラスを階層的に構成しており、他のオブジェクトとの通信は全て一番上位のクラスであるトップクラスで行う。各オブジェクトのクラス構成はツリー状になっているが、これは概念上の階層にもとづいており、実際のシステム構築の際には、継承などの機能はシステムのクラスを利用する時のみに用いている。

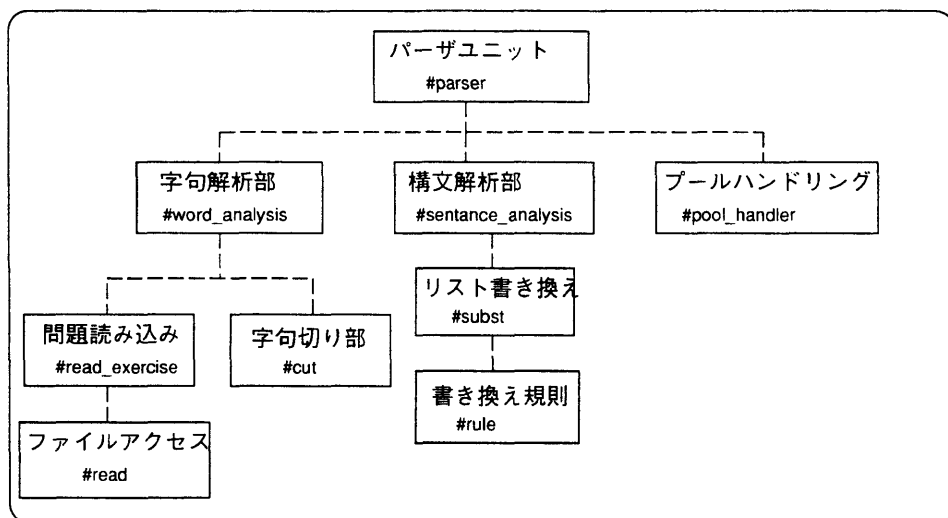


図 3: パーザ部オブジェクトのクラス構成

以下、各処理部について詳細を述べる。

### (1) パーザ部

パーザ部では、指定された番号の問題文を問題ファイル中から読み込み、字句解析部によって、字句切りを行う。また、構文解析部では字句解析後の問題文のリストより、後の処理を行いやすくなるためのリストの標準化を行う。

算数の文章題では、以下のようにアラビア数字を含まずに数を表わす場合がある。

- ひとつ、ふたつ
- ひとり、ふたり



この表現は別におかしいものではなく、むしろ日本語で表現する時にはこちらの方が自然に感じられる。従って、問題文の中ではこの表現を認め、構文解析部で以下のように標準化することによって処理を進める。

- ひとつ、ふたつ → 1つ、2つ
- ひとり、ふたり → 1にん、2にん

2文が1文にまとめられた複文などの展開処理などは、この構文解析部が受け持つべき部分であるが、現在のところあらかじめマニュアルで展開しておくこととする。

## (2) 問題解析部

問題解析部では、字句・構文解析部で得られたリスト形式の問題とあらかじめ定義した4種類の型の内部表現とユニフィケーションを行うことによって、問題のタイプを同定すると同時に内部表現を構築する。

またユニフィケーションを行なう順序はタイプ B,C が判断の決め手となる文が1文で良いのに対し、タイプ A は全文をチェックしなければいけないので、まずタイプ B,C であるかの判断をし、その後タイプ A であるかの判断を行う。タイプ D は常識により対象物を1つから2つに展開する例外処理を行なった後、ユニフィケーションを行わなければいけない場合があるので、タイプ A の後に判断を行う。

対象物、数詞、述語、キーワードの要素が足りない時、関係がおかしい時は、CESP のバックトラックにより次のタイプの判定へと移る。ユニフィケーションが正常に終了すると内部表現が得られ、ユニフィケーションに失敗すると解析に失敗したことになる。

文中から対象物や述語を抜き出すには、日本語の膠着性を利用する。

数と助数詞は数詞として対になっていることを利用し、述語は文章の最後に表われることを利用して判断する。算数問題の中には補助動詞を使用しているものもあるが、その場合はさらに1つ前の動詞を述語とみなす。対象物については日本語が持つ特徴より位置が不定なため、算数問題の文章中で位置が特定できない。そのため、対象物を見つけ出すのには名詞の後に続く助詞の特徴を利用する。算数問題に使われる助詞を目的別に分類したものを表1に示す。

表 1: 助詞の用法

は	対象物の後
を	
が	対象物が複数あるときの1番目の対象物の後
と	
の	対象物の様子を表すものの後

助詞を基準にして対象物を特定しようとすると、候補が複数個現れることがある。また、抜き出した対象物が1つでも必ずしも正しいものとは限らない。対象物の表現を代名詞で済ませていたり、省略されていたために別の物を対象物とみなしてしまう場合がある。このため、辞書の助数詞と動詞の各要素にはその助数詞、動詞を使用する名詞についての属性を持たせておき整合性を確認する処理を行う。この処理によって助数詞、動詞から抜き出してきた対象物が正しいものかどうかをチェックすることができる。また、主語の省略などにより対象物が文中に見つけられない時には、この知識を利

用して前の文章より対象物を抜き出してくることができる。この段階で問題文の対象物と助数詞、動詞の関係がおかしければ、対象物を見つけ出すことができず、解析は失敗する。

修飾子も問題によっては非常に重要な役目を果たす。修飾詞は必ず対象物の前に現れるという特徴を利用して抽出する。対象物が代名詞などに置き変わっている場合も同様にして行う。

タイプ A の問題では、質問文にキーワードが現れる場合があるが、これは問題解法を決定する上で重要な役割を果たすので、存在する場合にはこの処理部で同様に抽出する。このキーワードに関しても問題文の文脈に反するものがあった場合は解析は失敗する。

### (3) 推論・計算部

推論・計算部では、意味解析部で得られた内部表現より計算方法を決定し、実際の計算を行う。

計算方法の決定は問題解析部で作成された 3 文の内部表現をもとに行う。辞書の中の動詞には、それぞれ動詞の持つ意味から「くる +」、「いく -」の様に符号が与えてあり、問題文の 3 つの文のそれぞれの述語から、それぞれの文章の符号を決定し、この 3 文の符号の組み合わせによって計算方法を決定する。また、タイプ A の問題では条件文のみからでは計算方法が決定できず質問文のキーワードの情報が必要となる場合がある。したがってキーワードにも辞書の中に「みんなで +」のように符号を割り振っておき、そのような問題ではキーワードの符号もあわせて計算方法を決定する。この計算方法の決定を行なう符号の組み合わせは 9 通りある。表 2 に一覧を示す。表中の「？」はどちらともいえないことを表す。符号の与えられない文章があった場合は、他の文章に与えられた符号から推論を行う。

表 2: 計算方法の判断

条件文 1	条件文 2	質問文	計算方法
+	+	+	足し算
+	-	+	引き算
+	-	?	引き算
+	-	-	引き算
-	-	-	足し算
-	+	+	引き算
-	+	-	引き算
?	+	+	足し算
?	+	?	足し算

### (4) 文章生成部

文章生成部では、推論・計算部によって導き出された解答と対象物の変化の結果より、文章生成規則に従って答えの文章を生成する。

生成される解答の文章は、問題文より読みとった対象物、解答の数、助数詞、最終的に様子を表わす動詞を、各タイプの問題文毎に用意した解答テンプレートの各項目にユニフィケーションすることによって得られる。このとき、問題文中で対象物が違っていた場合はそれらの共通の属性を対象物として扱う。また解答によっては構文解析部で数詞の書き換えを行ったのと全く逆の処理を行わなくてはならない。問題のタイプごとの解答テンプレートは、次に示すものである。

1. タイプ A の問題  
(「修飾子」「助詞」)「対象物」「助詞」(「キーワード」「助詞」)「数詞」「動詞」
2. タイプ B の問題  
(「修飾子」「助詞」)「対象物」は「数詞」「動詞」
3. タイプ C の問題  
(「修飾子」「助詞」)「被対象物」が(「修飾子」「助詞」)「対象物」より「数詞」「動詞」
4. タイプ D の問題  
(「修飾子」「助詞」)「対象物」は「数詞」「動詞」

このテンプレートに対し、解答の内容にあわせて対象物や数詞、動詞等をユニフィケーションする。ここで「助詞」の場所には助詞の知識を利用して生成する文章にあう助詞を選ぶ。()内は入るものが無かったら省略する。

## (5) インターフェース部

インターフェース部は、入力用のウィンドウを作成しユーザから問題番号等の入力を受け付け、問題・導き出された解答を別に作成したウィンドウ上に表示する。

作成するウィンドウは、CESP が提供しているウィンドウ・ライブラリを利用する。具体的には、入力用のウィンドウは対話型ウィンドウ機能を提供するクラス `dialogbox` を継承することによってユーザとの対話入力を実現し、出力用のウィンドウはクラス `window` を継承し、さらにウィンドウ標準入出力クラスを継承することによって、ウィンドウに対して入出力(このシステムの場合出力のみ)を行う。また、出力用のウィンドウは入力用のウィンドウの子ウィンドウとして作成する。

## (6) 辞書アクセス部

辞書アクセス部は、各処理部に対して辞書へアクセスするメソッドを提供し、辞書の管理を行う。また辞書はテキストファイルで保有するものとし、システム内部ではハッシュを用いて管理を行う。そのため起動時に一度辞書ファイルより辞書ごとに一つずつのハッシュテーブルを作成する。

この処理部で利用するハッシュは、CESP が提供しているブール管理ライブラリのクラス `hash-index` を使用する。

辞書中の単語とのマッチング処理、属性情報の参照、動詞の符号参照などの辞書に対する処理は全てこの処理部で行ない、各処理部はこのモジュールへ辞書参照要求を出すことによって辞書に対する操作を行なう。

## (7) 辞書の構成

本研究では、辞書はサンプルの問題を解答するための必要最小限のもののみで構成する。また、それぞれの辞書は、テキストファイル上に一単語一行でそれぞれの単語と、それぞれの単語が持つ属性などを空白記号を区切り文字として記述する。

辞書の種類を以下に示す。

名詞辞書	：名詞・その単語が持つ属性の知識
助詞辞書	：助詞・助詞の分類結果
動詞辞書	：動詞・その動作を行う単語の持つ属性・計算方法を決定する符号

**助数詞辞書** : 助数詞・その助数詞を用いる単語の持つ属性・その助数詞の変化形

**キーワード辞書** : キーワード・計算方法を決定する符号

この他の、修飾詞、代名詞等の辞書には、単に一行に一つの単語を書く。

推論の際に必要な常識と呼ばれる知識に関しても、辞書の一つとして持つ。この辞書には、同一の事象を指している単語を1行に列記する。

### 3.2 実行結果

小学校1年生の算数の問題集2冊の中から条件に合うものを選び出し、このシステムで解析してみた。その結果、タイプAの問題49題、タイプB8題、タイプC3題、タイプD12題の合計72題について解析可能であった。

例えばタイプCの代表的な問題

おとこのこが 34 にん いる。  
 おんなのこが 20 にん いる。  
 おとこのこは なんにん おおいでしょう。

に対して、システムは以下の解答をウィンドウ上に出力する。

おとこのこはおんなのこより 14 にんおおい。

しかし、2章で述べた通り対象物が問題中で変化してしまうような問題や、次のような対象物が文章から同定できないような問題は解析できない。

つくしとりで ぼくは 20 ぽん とる。  
 ねえさんは ぼくより 7 ほん おおく とる。  
 ねえさんは なんぽん とったでしょう。

この問題では、「つくしとり」という言葉で表されている状況や行為を理解しないと、対象物が「つくし」であることがわからない。また、この問題における比較の対象は「ねえさんの(とった)つくし」と「ぼくの(とった)つくし」であるが、「とる」という述語だけでは判断できず、「つくしとり」と合わせてはじめて理解することができるようになる。

## 4 考察と展望

以上、小学校低学年の算数文章題のうち足し算・引き算の解析について述べてきた。前章で述べたように、問題文の形式に制限はあるものの、ある程度の問題は詳細な構文解析等を行なわなくとも、意味理解の点から必要な情報の抽出と、問題の分類による解析が可能なが判った。

今後知識表現を見直していく際に、現在のところシステムの構築のみに利用したオブジェクト指向の考えをシステムの知識表現の構成に拡張する必要がある。これにより、現在は属性という概念は利用しているもののフラットな概念でしかとらえられていなかったシステム内知識をより自然に表現することが可能になるのではないかと考えられる。また、これにより2章で述べた常識などの知識、「こども」が「おとこのこ」、「おんなのこ」に排他的に分かれるという知識などは現在のような例外的処理でなく、より自然に、よりの確に処理できるのではないかと考える。

計算機により算数問題を解析するには、基本的には文章を解析した結果を記号論的に形式化し、形式的推論の枠組に合わせて推論を組み立てていくことになる。その意味では、問題の「類型分類」と「類型のパターン化」である。今回、小学校低学年の足し算・引き算の問題を4つのタイプに分類することにより、ある程度簡単に解答を導き出せることが判った。ただし、この場合の分類とは、教科書や参考書にあるような分類ではなく、形式化された知識表現の観点からの類型化である。しかし、問題の対象範囲を掛け算・割り算を含む複合演算にまで拡大していこうと考えた時、現在の分類、内部表現の延長では困難なことが判った。そこで、今後、内部表現の見直し、推論方法の見直し等を行ない、問題の対象範囲を拡大していくことが必要であると思われる。

また、問題文の形式も、もっと柔軟に対応するため、字句解析、構文解析を充実させる必要もある。そして、これらの各処理部が互いに情報をフィードバックしながら処理を進めることによって、もっと柔軟な内部表現が可能となり、対象範囲の拡大にも繋がると思われる。

今後このシステムの解析可能な問題の対象範囲を上げ、最終的にはつるかめ算に代表されるような複合演算の処理を可能とするため、まず掛け算・割り算の問題の処理を考える。

小学校低学年の算数問題に使われる掛け算・割り算の文章題の多くは、九九を理解するための基本的な問題であり、1回の計算で答が求まるものである。例えば、

1 そく 500 えんの くつしたがあります。  
5 そく かいました。  
なんえん はらえば いいでしょう。

この問題をこれまでの4つの分類と同じように、意味的な文構造から考えると、1文目に特徴が見られ、この文で1つの対象物に対する2つの数の関係が述べられている。

また、次のような問題では、2つの対象物の数の関係が1文目で述べられている。

ジュースが はこに 6ぼんずつ はいっています。  
はこは 3はこ あります。  
ジュースは みんなで なんぼん あるでしょう。

掛け算・割り算の問題ではこのように、2つの数の関係から解答を導くための計算式を導出することが求められる。そこで、これまでに述べてきた4つのタイプにはなかった、2つの数の関係を表現する引数を作り、掛け算の問題を内部表現することを試みる必要がある。問題の分類の基準は、とりあえず数詞が1文に2つ現われる問題をこのタイプの問題(掛け算)とする。

この分類と内部表現および推論規則を前章のシステムに組み込むことによって、実際に単純な掛け算の文章題はいくつか解析できる。しかし割り算の問題も基本的には掛け算と同じ構造をしているために、この分類では掛け算・割り算を区別することが極めて困難なことが予想できる。足し算・引き算の区別を決定するには、述語を利用すれば3文の述語から殆どの問題で区別が可能であるが、掛け算・割り算の区別は述語からでは殆ど不可能である。それは基本的に掛け算・割り算の計算方法は先に述べた通り、文章中に現われる2つの数字の関係によって決定され、しかも、その関係は述語を見れば判るというものではない。このような関係を解析するためには、問題の文章全体の解釈、つまり述語といくつかのキーワードにだけ頼った解析でなく、文章全体の文脈理解を行なわなければいけない。これは今後、いくつかの演算が交じった複合演算を求める文章題に対して解析を行なおうとする際にも必要となると思われる。

## 参考文献

- [1] 高木 朗, 伊東幸宏:「自然言語の処理」, 丸善株式会社 (1987)

- [2] 後藤滋樹：「PROLOG 入門」，サイエンス社 (1984)
- [3] 溝口文雄 監修，溝口文雄・武田正之・畠見達夫・溝口理一郎：「Prolog とその応用 2」，総研出版 (1985)
- [4] L.Sterling・E.Shapiro 著，松田利夫 訳：「Prolog の技芸」，共立出版 (1988)
- [5] W.F.Clocksini/C.S.Mellish 著，中村克彦 訳：「Prolog プログラミング [改訂第 3 版]」，マイクロソフトウェア (1988)
- [6] 中島秀之：「Prolog」，産業図書 (1983)
- [7] 「新改訂さんすう 1 ねん」，啓林館
- [8] 「新改訂さんすう 2 年上」，啓林館
- [9] 「算数ミニドリル/算数文章題小学校 1 年」，教育書籍
- [10] 「小学 1 年文章題 5 分間トレーニング」，教学研究社
- [11] 近山 隆・中澤 修・内田洋一・佐藤泰典・田中吉廣・実近憲明，「論理型オブジェクト指向言語 ESP」，共立出版 bit 連載.vol22 No1-12 (1990)
- [12] 小倉久和・金田美佳：「コンピュータによる算数問題の理解と表現」，福井大学情報処理センター NETWORK, vol.6, No.2,57-70 (1992)
- [13] 笹田 国博：「CESP による算数問題文の理解と解析」，福井大学工学部情報工学科 卒業論文，(1993)
- [14] 笹田 国博，金田 美佳，小高 知宏，小倉 久和：「算数問題文の知識表現と理解システム」，平成 5 年度電気関係学会北陸支部連合大会講演論文集 p349, (1993)
- [15] 小倉 久和，笹田 国博：「オブジェクト指向型言語 CESP による算数問題文の理解と推論の試み」，情報処理学会第 46 回全国大会講演論文集 (3) pp.195-196, (1993)